

UDC: 004.89

Approaches for image processing in the decision support system of the center for automated recording of administrative offenses of the road traffic

R. N. Minnikhanov^{1,2,a}, I. V. Anikin^{2,b}, M. V. Dagaeva^{1,2,b},
T. I. Asliamov^{1,2,c}, T. E. Bolshakov^{1,2,d}

¹“Road Safety” State Company,
5 Orenburgskij trakt, Kazan, 420059, Russia

²Kazan National Research Technical University named after A. N. Tupolev — KAI,
10 Karl Marx st., Kazan, 420111, Russia

E-mail: ^aRifkat16@gmail.com, ^bAnikinigor777@mail.ru, ^cDagaevam@rambler.ru,
^dAslyamovt@gmail.com, ^eBear_333@mail.ru

*Received 14.09.2020, after completion — 29.01.2021.
Accepted for publication 01.02.2021.*

We suggested some approaches for solving image processing tasks in the decision support system (DSS) of the Center for Automated Recording of Administrative Offenses of the Road Traffic (CARAO). The main task of this system is to assist the operator in obtaining accurate information about the vehicle registration plate and the vehicle brand/model based on images obtained from the photo and video recording systems. We suggested the approach for vehicle registration plate recognition and brand/model classification on the images based on modern neural network models. LPRNet neural network model supplemented by Spatial Transformer Layer was used to recognize the vehicle registration plate. The ResNeXt-101-32x8d neural network model was used to classify for vehicle brand/model. We suggested the approach to construct the training set for the neural network of vehicle registration plate recognition. The approach is based on computer vision methods and machine learning algorithms. The SIFT algorithm was used to detect and describe local features on images with the vehicle registration plate. DBSCAN clustering was used to detect and delete outliers in such local features. The accuracy of vehicle registration plate recognition was 96% on the testing set. We suggested the approach to improve the efficiency of using the ResNeXt-101-32x8d model at additional training and classification stages. The approach is based on the new architecture of convolutional neural networks with “freezing” weight coefficients of convolutional layers, an additional convolutional layer for parallelizing the classification process, and a set of binary classifiers at the output. This approach significantly reduced the time of additional training of neural network when new vehicle brand/model classification was needed. The final accuracy of vehicle brand/model classification was 99% on the testing set. The proposed approaches were tested and implemented in the DSS of the CARAO of the Republic of Tatarstan.

Keywords: decision-support system, video image, computer vision, neural networks

Citation: *Computer Research and Modeling*, 2021, vol. 13, no. 2, pp. 405–415 (Russian).

© 2021 Rifkat N. Minnikhanov, Igor. V. Anikin, Maria V. Dagaeva,
Timur I. Asliamov, Tikhon. E. Bolshakov

This work is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/3.0/>
or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

УДК: 004.89

Подходы к обработке изображений в системе поддержки принятия решений центра автоматизированной фиксации административных правонарушений дорожного движения

Р. Н. Минниханов^{1,2,a}, И. В. Аникин^{2,b}, М. В. Дагаева^{1,2,c},
Т. И. Аслямов^{1,2,d}, Т. Е. Большаков^{1,2,e}

¹ ГБУ «Безопасность дорожного движения», Россия, 420059, г. Казань, Оренбургский тракт, д. 5

² Казанский национальный исследовательский технический университет им. А. Н. Туполева — КАИ,
Россия, 420111, г. Казань, ул. К. Маркса, д. 10

E-mail: ^a Rifkat16@gmail.com, ^b Anikinigor777@mail.ru, ^c Dagaevam@rambler.ru,
^d Aslyamovt@gmail.com, ^e Bear_333@mail.ru

Получено 14.09.2020, после доработки — 29.01.2021.

Принято к публикации 01.02.2021.

В статье предлагается ряд подходов к обработке изображений в системе поддержки принятия решений (СППР) центра автоматизированной фиксации административных правонарушений дорожного движения (ЦАФАП). Основной задачей данной СППР является помощь человеку-оператору в получении точной информации о государственном регистрационном знаке (ГРЗ) и модели транспортного средства (ТС) на основании изображений, полученных с комплексов фотовидеофиксации (ФВФ). В статье предложены подходы к распознаванию ГРЗ и марки/модели ТС на изображении, основанные на современных нейросетевых моделях. Для распознавания ГРЗ использована нейросетевая модель LPRNet с дополнительно введенным Spatial Transformer Layer для предобработки изображения. Для автоматического определения марки/модели ТС на изображении использована нейросетевая архитектура ResNeXt-101-32x8d. Предложен подход к формированию обучающей выборки для нейросетевой модели распознавания ГРЗ, основанный на методах компьютерного зрения и алгоритмах машинного обучения. В данном подходе использован алгоритм SIFT для нахождения ключевых точек изображения с ГРЗ и вычисления их дескрипторов, а для удаления точек-выбросов использован алгоритм DBSCAN. Точность распознавания ГРЗ на тестовой выборке составила 96 %. Предложен подход к повышению производительности процедур дообучения и распознавания марки/модели ТС, основанный на использовании новой архитектуры сверточной нейронной сети с «заморозкой» весовых коэффициентов сверточных слоев, дополнительным сверточным слоем распараллеливания процесса классификации и множеством бинарных классификаторов на выходе. Применение новой архитектуры позволило на несколько порядков уменьшить время дообучения нейросетевой модели распознавания марки/модели ТС с итоговой точностью классификации, близкой к 99 %. Предложенные подходы были апробированы и внедрены в СППР ЦАФАП Республики Татарстан.

Ключевые слова: система поддержки принятия решений, изображение, компьютерное зрение, нейронные сети

© 2021 Рифкат Нургалиевич Минниханов, Игорь Вячеславович Аникин, Мария Витальевна Дагаева,
Тимур Ильдарович Аслямов, Тихон Евгеньевич Большаков

Статья доступна по лицензии Creative Commons Attribution-NoDerivs 3.0 Unported License.

Чтобы получить текст лицензии, посетите веб-сайт <http://creativecommons.org/licenses/by-nd/3.0/>
или отправьте письмо в Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1. Introduction

The main functions of modern CARAO involve automatic detection of traffic rules violation by vehicles (including high speed, stopping and parking rules, etc.), as well as registration of corresponding acts. Video enforcement cameras are widely used now for the detection of such violations. Their usage in the city of Kazan has reduced the number of fatalities in car accidents by about 70% per year, starting from 2012.

To detect traffic rules violations, video enforcement complexes collect and process a great number of images from the streets, recognize vehicle registration plates and detect vehicle speed. After that, the CARAO operator manually solves the following tasks to make an official decision about rules violation of the rules:

- 1) confirmation of recognized vehicle registration plate number;
- 2) detection of vehicle brand and model and comprising them with the data obtained from the traffic police database.

The CARAO operator manually solves the above-mentioned tasks, which significantly complicates the process of decision making, increases labor costs, and the number of errors caused by the human factor. Therefore, developing an automatic decision-making system, which assists the CARAO operator, is important. This article is devoted to this problem.

In Section 2 we suggested an approach for vehicle registration plate recognition based on neural networks. In Section 3 we suggested an approach to construct a training set for the neural network of vehicle registration plate recognition. In Section 4 we suggested an effective approach to classify vehicle brand/model on the images. In Section 5 we make experiments and present evaluation results.

2. The approach to recognize a vehicle registration plate

Nowadays, artificial neural networks are widely used to recognize vehicle registration plates [Davix et al., 2020; Omar et al., 2020; Nayak et al., 2020; Izidio et al., 2020; Dias et al., 2019; Obukhov et al., 2016]. However, some of the existing methods are not accurate, not fast enough, have recognition errors on noisy images, or images with a wrong tilt angle. Moreover, many methods are based on preliminary segmentation of a vehicle registration plate to separate characters and recognize them independently, which may increase the number of errors.

In this paper, we suggest an approach to recognize a vehicle registration plate based on the following steps:

- 1) Detection of the vehicle registration plate on the image. For this, we used the `fasterrcnn_resnet50_fpn` neural network model, which showed high accuracy and performance for object detection tasks on the images [Tian, Chen, 2019].

- 2) Symbols recognition of the detected vehicle registration plates. We used the LPRNet neural network model [Zherzdev, Gruzdev, 2018], which works without preliminary symbolic segmentation of the vehicle registration plate. The architecture of the LPRNet neural network model is shown in Table 1. To obtain more accurate recognition results, the vehicle registration plate image was preliminarily transformed using the Spatial Transformer Layer [Jaderberg et al., 2016], which is jointly trained with LPRNet. The output of the LPRNet neural network model is the set of recognized symbols of the vehicle registration plate (Figure 1).

Despite the fact that the above two steps are quite clear and used novel but well-described neural network models, the main problem which arises is the preparation of a good training set for `fasterrcnn_resnet50_fpn`, LPRNet with Spatial Transformer Layer neural network models. The approach to construct such a training set is proposed below.

Table 1. LPRNet Neural Network Model Architecture

Layer Type	Parameters
Input	94x24 pixels RGB image
Convolution	#64 3x3 stride 1
MaxPooling	#64 3x3 stride 1
Small basic block	#128 3x3 stride 1
MaxPooling	#64 3x3 stride (2, 1)
Small basic block	#256 3x3 stride 1
Small basic block	#256 3x3 stride 1
MaxPooling	#64 3x3 stride (2, 1)
Dropout	0.5 ratio
Convolution	#256 4x3 stride 1
Dropout	0.5 ratio
Convolution	#class_number 1x13 stride 1

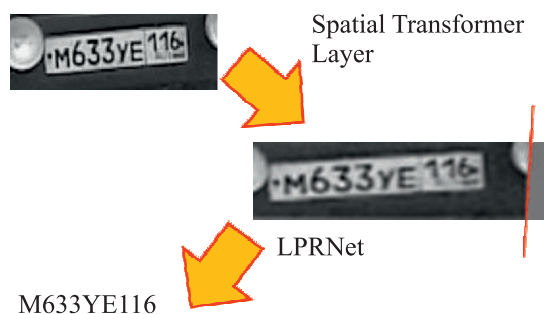


Fig. 1. Vehicle registration plate recognition using LPRNet and Spatial Transformer Layer

3. Construction of the training set for the vehicle registration plate recognition task

The training set for fasterrcnn resnet50 fpn neural network model has to include:

- vehicle images obtained from video enforcement cameras;
- coordinates of the vehicle registration plate on the corresponding vehicle images.

To create such a training set, we process 157470 vehicle images from video enforcement cameras and corresponding vehicle registration plate images by the following steps (Figure 2).

Step 1. Find a set of key points on the vehicle image and calculate their descriptors.

The key points are special points of the image that are very likely to be detected in other images of the class. The descriptor is the identifier of the key point that separates it from the remaining group. Using descriptors allows ensuring invariance of the procedure for various transformations.

In this step, we used the well-known SIFT algorithm to determine the image's key points and their descriptors [Lowe, 1999]. 157470 reference images with vehicle registration plates (Fig. 3) are used to extract the key points.

The SIFT algorithm uses key points as local extrema of difference between different Gaussians [Shapiro, Stockman, 2001] of the following type:

$$D(x,y,\sigma) = L(x,y,k,\sigma) - L(x,y,k,\sigma).$$

The expression $L(x,y,k,\sigma)$ is a convolution of the original image $I(x,y)$ with the Gauss filter $G(x,y,k,\sigma)$ with the degree of blurring σ and the scale k . The convolution images are grouped by octaves consisting of images blurred by a Gaussian filter with a different blurring radius. Each subsequent octave corresponds to doubling the value of σ . The number s of intervals in the octave is selected, then

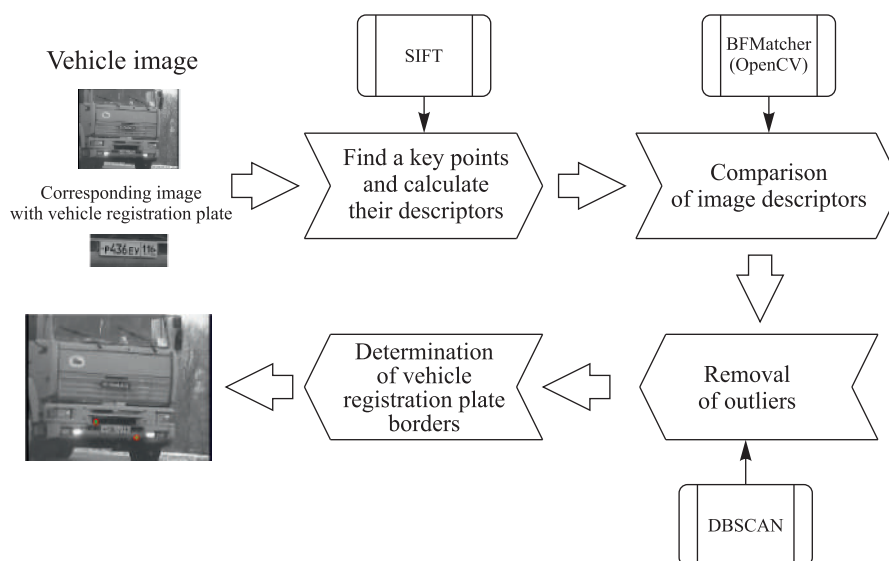


Fig. 2. Creation of the training set for fasterrcnn_resnet50_fpn neural network model



Fig. 3. Vehicle registration plate image from the set of reference images

$(s + 3)$ images with $k = 2^{1/s}$ for each octave are built. The original image of the next octave is obtained by sampling the image from the previous octave (every second row and every second column is taken). The Gaussian difference from adjacent Gaussian blurred images in the octave is then calculated and local extrema are found. The key points obtained are filtered by removal of points with subpixel accuracy, low contrast, and border objects. Finally, orientation histograms around key points are analyzed to construct the descriptors.

The SIFT method is invariant under image rotation, scale, and lighting changes. It is also resistant to noise and computationally efficient. Figure 4 shows an example of a vehicle image obtained from the video enforcement camera and the results of applying the SIFT algorithm to it and its vehicle registration plate.

Step 2. Comparison of image descriptors.

The key point descriptor is a unit length vector, which is produced from orientation histogram entries values of the corresponding key point. The unit length was used to be invariant under image contrast changes. Thresholding of the vector values by the value 0.2 was used to be invariant under non-linear illumination changes [Lowe, 1999].

Comparison of such descriptors is used to find key point elements of the one image on another one. It becomes possible to find more easily any specific objects on a large image, in the present case, to find a specified vehicle registration plate on the corresponding vehicle image.

To compare two key points descriptors, we used the method BFMatcher of the OpenCV library. This method takes a certain descriptor and is matched with all others, and the closest descriptor is returned. Euclidean distance was used for such matching. Figure 5 shows the result of this step in the form of projection of key points to the original image.

Step 3. Removing outliers. As we can see in Figure 5, most of the matching key points are dislocated in the area of the vehicle registration plate image. However, there are some outliers localized mainly at the image borders. The well-known DBSCAN density clustering algorithm was used to filter such outliers [Ester et al., 1996]. We selected the cluster with the maximum number of points on the image as the vehicle registration plate cluster. Figure 6 shows the result of removing outlier points.



Fig. 4. Picture captured by the video enforcement camera and their key points



Fig. 5. Comparison of image descriptors obtained using the SIFT algorithm



Fig. 6. Removal of outliers

Step 4. Determination of vehicle registration plate borders

Borders of the vehicle registration plate image were obtained by analysis of the remaining key points. The upper left point coordinates (X_1, Y_1) of the vehicle registration plate is the mean difference by x and y between the key points of the vehicle registration plate image and corresponding key points of the vehicle image. The right bottom point coordinates (X_2, Y_2) were determined by summarizing the coordinates (X_1, Y_1) with the width and length of the vehicle registration plate image obtained from the video enforcement camera.

We used the training set with 157 470 vehicle images and corresponding coordinates of the vehicle registration plate (extracted in Steps 1–4) to train the fasterrcnn_resnet50_fpn neural network model.

We also used the training set with 157 470 vehicle registration plate images and correct vehicle numbers for training the LPRNet neural network model.

4. Classification of vehicle brand/models

There are many research articles devoted to vehicle classification on the images using deep neural networks [Bhujbal, Mane, 2020; Chen et al., 2019; Liu et al., 2019; Hicham et al., 2018; Arinaldi et al., 2018]. However, most of them are devoted to vehicle type identification without specifying a brand/model. Also, it is a computationally complex task to constantly train or retrain the existing deep neural network to classify a new constantly expanding number of vehicle models.

Table 2 presents the authors' comparative analysis of different neural network architectures to solve classification problems for 19 vehicle models on 13170 images, performed within one epoch on the PC with Intel Core i5 8400H PC, Nvidia Geforce 1050. As a result, the ResNeXt-101-32x8d neural network architecture was chosen to solve this problem, which showed the best accuracy and acceptable performance.

Table 2. Comparison of different neural network classification architectures

Neural Network Architecture	Accuracy	Average recognition time (sec)	Training time
resnext101_32x8d_X11Epoch_SGD	0.69	0.05	33m 57s
resnext50_32x4d	0.65	0.024	14m 45s
densenet201	0.64	0.077	16m 47s
densenet121	0.62	0.044	9m 52s
resnet101	0.60	0.032	14m 25s
resnet34	0.60	0.013	5m 32s
resnet50	0.59	0.016	8m 33s
mobilenet_v2	0.57	0.016	5m 8s
densenet169	0.57	0.063	12m 56s
Googlenet	0.55	0.023	4m 36s
inception_v3	0.53	0.031	8m 18s

The image of the vehicle obtained from the video enforcement camera is fed to the input of the ResNeXt-101-32x8d neural network model, the vehicle brand and model (459 classes) are produced at the output (Figure 7). We used 5715161 labeled images of the vehicles obtained from the video enforcement cameras as a training set. Initially, we used ResNeXt-101 with the pre-trained weights on the imagenet data set, and then we trained it on the vehicles data set.

The main problem of the neural network model shown in Figure 7 is its learning rate, which makes it difficult to retrain this model when new vehicle brands/models appear. It takes about 10 days (6 epochs on TESLA V100 16G) on the training set with full vehicle brands/models. It is completely unacceptable when we need to add a new brand/model in our classification process. To solve this

problem we suggested a classification model based on the new architecture of a convolutional neural network. This architecture is based on 459 binary classifiers trained to recognize a particular brand/model of the vehicle on the “frozen” values of convolution layers, obtained during the initial training of ResNeXt. Figure 8 shows the neural network architecture with binary classifiers.

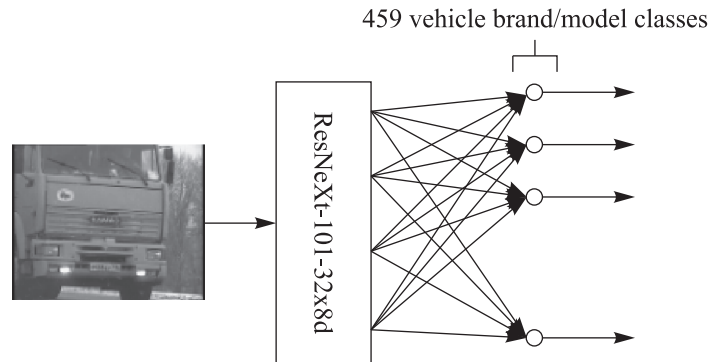


Fig. 7. Initial neural network architecture for vehicle brand/model classification

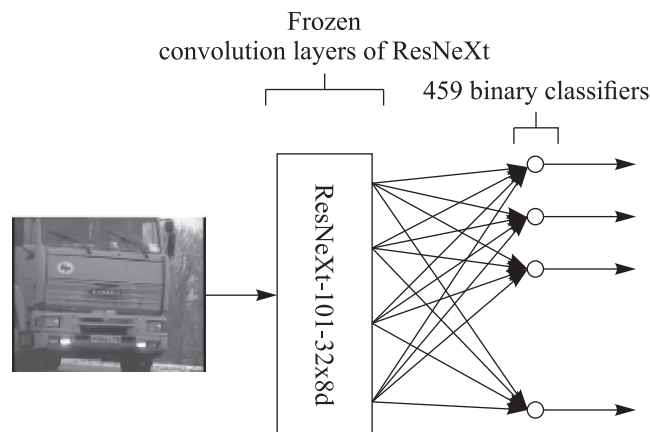


Fig. 8. Neural network architecture with binary classifiers

After the training and “freezing” of convolution layers for ResNeXt model, every binary classifier is trained on the training set consisting of positive and negative examples of vehicle models. The number of negative examples in the training set was taken to be equal to the number of positive examples. Such an approach can significantly reduce the training time of the new binary classifier to recognize new models of vehicles because additional training of “frozen” layers is not needed in this case.

The bottleneck of the architecture, shown in Figure 8, is the need to obtain the outputs of all 459 binary classifiers for image classification. This fact significantly reduces the final performance of the system with no parallel computations. To solve this problem, we propose to take away all weights of pre-trained 459 binary classifiers into one convolution layer. We can get the output of this convolution layer in parallel processing mode.

Figure 9 shows the architecture of the final neural network with an additional convolution layer. The number M is the doubled dimension of the output vector of “frozen” layers of ResNeXt. The number $N = 459$ is the number of convolution layer channels corresponding to the number of binary classifiers. The output of the additional convolution layer is two vectors with dimension N , whose elements reflect the degree of the possibility that source data belongs to a certain class. We used the following parameters in the training step: 1) Optimizer: SSG; 2) Learning rate: 0.001; 3) Moment: 0.9; 4) Augmentation: horizontal image reflection, image scaling, cutting an arbitrary 224×224 piece.

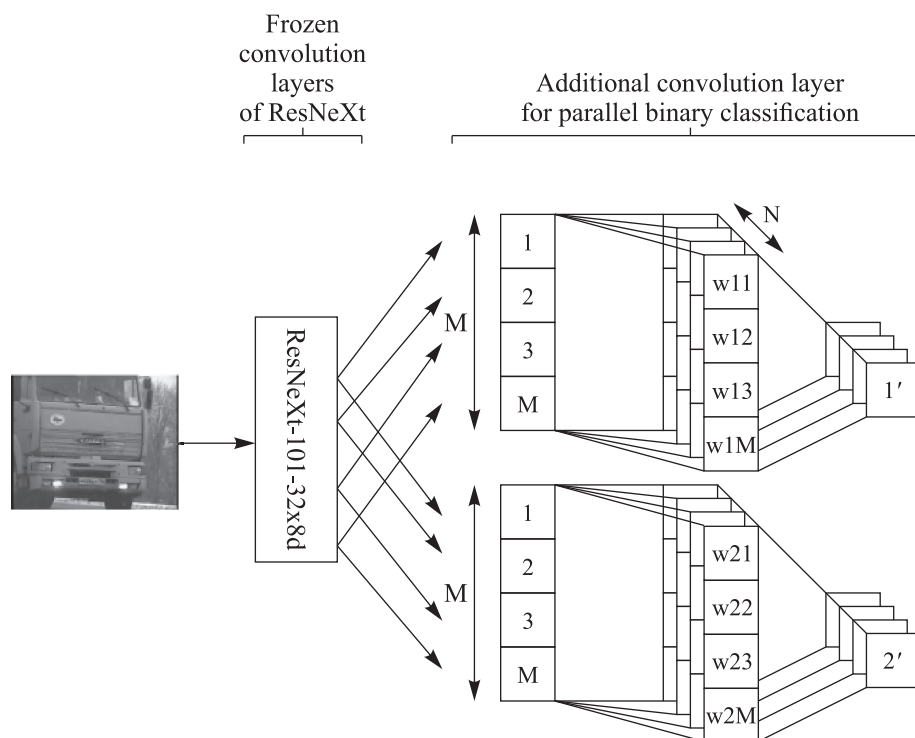


Fig. 9. The final architecture of the neural network for vehicle brand/model classification

5. Evaluation results

To evaluate our results, all suggested neural network architectures were trained on the training sets with 459 vehicle brand/model images and corresponding vehicle registration plates. The recognition accuracy of the vehicle registration plate was 96% on the testing set. We also got vehicle brand/model recognition accuracy on binary classifiers as 87% in the “one against all” mode. Also, we got vehicle brand/model recognition accuracy close to 99% if the vehicle can belong to several classes (according to the outputs of the binary classifiers).

Table 2 shows the training performance for one binary classifier (on TESLA V100 16G), which is the part of the model in Fig. 9. This performance is much better than the performance of the original neural network model, shown in Fig. 7 (10 days for 5715161 images on TESLA V100 16G). Consequently, the model shown in Fig. 9 can be effectively used for training the classifier when new vehicle models appear.

Table 2. Performance of binary classifiers

Training set size	Training time (sec)
2500	116
4000	255
8100	419
36000	2793
74000	7042

Also, the use of neural network architecture shown in Fig. 9 allows reducing the image classification time from 330 ms to 5 ms compared with the neural network architecture, shown in Fig. 8. The experiments were provided for 459 vehicle classes on PC with Intel Core i5 8400H and Nvidia Geforce 1050. Thus, using the final neural network architecture allows getting a significant effect both in training and classification time.

Figure 10 shows an example of vehicle brand/model, registration plate zone, and number recognition on the image obtained from the video enforcement camera.

All proposed approaches were integrated into the decision support system of the Center for Automated Recording of Administrative Offenses of the Road Traffic in the city of Kazan. This decision support system is working as a web-service with images obtained from the video enforcement camera. It gives a result in the following form:

```
{"brand": "MAZDA", "coord": "[(205.80414, 328.4791), (436.0373, 412.45908)]",
"id": "11111.jpg", "model": "3", "plate": "C234AC102", "probability": 99.99, "veh_type": "B"}
```

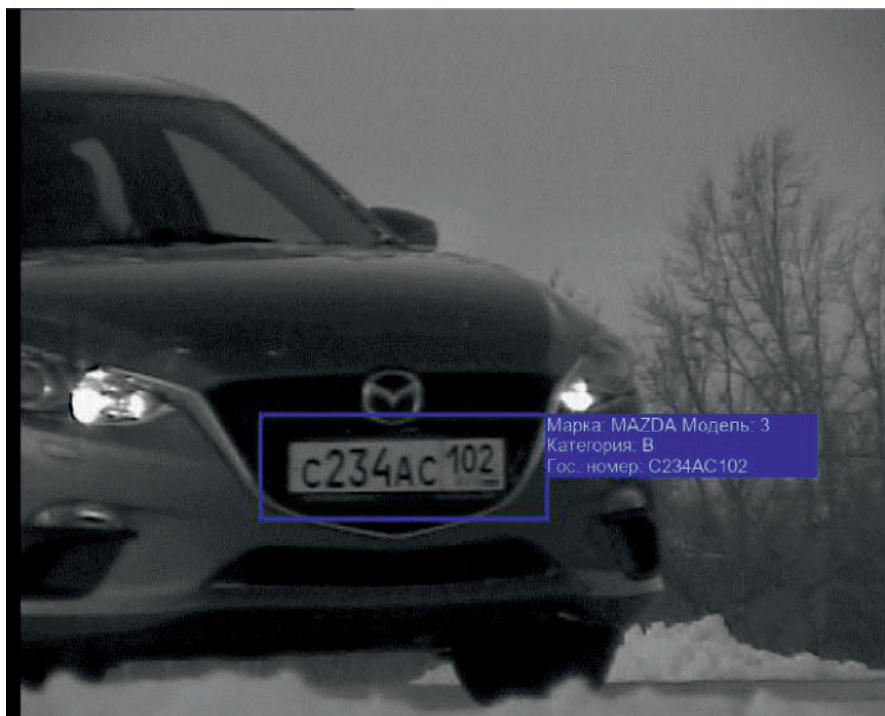


Fig. 10. An example of vehicle brand/model and registration plate recognition

6. Conclusion

We have solved the task of developing the decision support system for the operator of the Center for Automated Recording of Administrative Offenses of the Road Traffic. This system helps the operator to confirm the recognized vehicle registration plate number and to detect its brand and model. We used a neural network approach for solving these tasks. We also gave some suggestions for an effective usage of proposed neural network models.

We suggested an approach for the construction of the training sets for fasterrcnn_resnet50_fpn and LPRNet neural networks models for the task of detection and recognition of vehicle registration plate on images. The final accuracy of vehicle registration plate recognition was 96%.

We suggested an approach for increasing the effectiveness of using ResNeXt-101-32x8d neural network architecture for the task of vehicle brand/model classification. This approach is based on new neural network architecture with “freezing” convolution layer weights and additional convolution layer for binary classifiers. The use of this approach significantly reduced the time of additional training of the original ResNeXt-101-32x8d architecture when new vehicle models appeared. We also reduced the vehicle model classification time by a factor of more than 60. The final classification accuracy of the proposed neural network model is close to 99%.

We used a decision-support system in the Center for Automated Recording of Administrative Offenses of the Road Traffic in the city of Kazan and significantly reduced the operator workload.

References

- Arinaldi A., Pradana J. A., Gurusinga A. A.* Detection and classification of vehicles for traffic video analytics // *Procedia Computer Science*. — 2018. — Vol. 144. — P. 259–268.
- Bhujbal A., Mane D. T.* Vehicle Type Classification Using Deep Learning // *Advances in Intelligent Systems and Computing*. — 2020. — Vol. 1118. — P. 279–290.
- Chen L., Sun P.-Y., Jia Y., Sinnott R. O.* Identification and classification of trucks and trailers on the road network through deep learning // *BDCAT 2019 – Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*. — December 2019. — P. 117–126.
- Davix X. A., Christopher C. S., Judson D.* Detection of the vehicle license plate using a kernel density with default search radius algorithm filter // *Optik*. — 2020. — Vol. 218.
- Dias C., Jagetiya A., Chaurasia S.* Anonymous vehicle detection for secure campuses: A framework for license plate recognition using deep learning // *2019 2nd International Conference on Intelligent Communication and Computational Techniques, ICCT 2019*. — 2019. — P. 79–82.
- Ester M., Kriegel H.-P., Sander J., Xu X.* A density-based algorithm for discovering clusters in large spatial databases with noise // *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. — 1996. — P. 226–231.
- Hicham B., Ahmed A., Mohammed M.* Vehicle Type Classification Using Convolutional Neural Network // *Colloquium in Information Science and Technology, CIST Volume*. — 2018.
- Izidio D. M. F., Ferreira A. P. A., Medeiros H. R., Barros E. N. S.* An embedded automatic license plate recognition system using deep learning // *Design Automation for Embedded Systems*. — 2020. — Vol. 24, Iss. 1. — P. 23–43.
- Jaderberg M., Simonyan K., Zisserman A., Kavukcuoglu K.* Spatial Transformer Networks // <https://arxiv.org/abs/1506.02025>. 2016.
- Liu J., Shen F., Wei M., Zhang V., Zeng V.* A Large-Scale Benchmark for Vehicle Logo Recognition // *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*. — 2019.
- Lowe D. G.* Object recognition from local scale-invariant features // *Proceedings of the International Conference on Computer Vision*. — 1999. — Vol. 2. — P. 1150–1157.
- Nayak V., Holla S. P., Akshayakumar K. M., Gururaj C.* Automatic number plate recognition // *International Journal of Advanced Trends in Computer Science and Engineering*. — 2020. — Vol. 9, Iss. 3. — P. 3783–3787.
- Obukhov A., Lyasheva S., Shlymovich M.* The methods of automatic license plate recognition // *Vestnik Chuvashskogo universiteta*. — 2016. — No. 3. — P. 201–208.
- Omar N., Abdulazeez A. M., Sengur A., Al-Ali S. G. S.* Cascaded deep learning-based efficient approach for license plate detection and recognition // *Expert Systems with Applications*. — 2020. — Vol. 149.
- Shapiro L. G., Stockman G. C.* *Computer Vision*. — Prentice Hall, 2001. — P. 137–150.
- Tian X., Chen C.* Modulation Pattern Recognition Based on Resnet50 Neural Network // *2nd IEEE International Conference on Information Communication and Signal Processing, ICICSP*. — 2019.
- Zherzdev S., Gruzdev A.* LPRNet: License Plate Recognition via Deep Neural Networks // <https://arxiv.org/abs/1806.10447/>. 2018.

